. 478
959

Wallach

ol. 73
lo. 8

Psychological
Monographs
General and Applied

The Influence of Classification Requirements on Gradients of Response

By

Michael A. Wallach

Harvard University

Price $1.00

Edited by Norman L. Munn
Published by the American Psychological Association, Inc.

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

NORMAN L. MUNN, Editor

Department of Psychology, Bowdoin College
Brunswick, Maine

Consulting Editors

ANNE ANASTASI
FRANK A. BEACH
ARNOLD M. BINDER
W. J. BROGDEN
ROBERT R. BUSH
JOHN F. DASHIELL
JAMES J. GIBSON
D. O. HEBB
EDNA HEIDBREDER
FRANCIS W. IRWIN
JAMES J. JENKINS
HAROLD E. JONES
DANIEL KATZ
BOYD MCCANDLESS
DONALD W. MACKINNON
QUINN MCNEMAR
LORRIN A. RIGGS
CARL R. ROGERS
RICHARD L. SOLOMON
ROSS STAGNER

Manuscripts and correspondence on editorial matters should be sent to the Editor. *Psychological Monographs* publishes comprehensive experimental investigations and programmatic studies which do not lend themselves to adequate presentation as journal articles. Major space is given to the author's original contribution; introductory and bibliographic materials, as well as statistical tables and graphs, must be kept within reasonable bounds. Tables, graphs, and appendix materials which deal with detail not essential to adequate presentation of the findings may be made available through the American Documentation Institute—for details of this procedure, see the APA *Publication Manual.* Preparation of manuscripts for publication as monographs should follow the procedure given in the APA *Publication Manual.* Publication in *Psychological Monographs* is free of cost to the author, except in cases where early publication is requested or author's alterations are made in galley proofs.

ARTHUR C. HOFFMAN, Managing Ed.; HELEN ORR, Promotion Mgr.; SADIE J. DOYLE, Editorial Asst.

Correspondence on business matters should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St., N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

COPYRIGHT, 1959, BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Vol. 73, No. 8 Whole No. 478, 1959

Psychological Monographs: General and Applied

THE INFLUENCE OF CLASSIFICATION REQUIREMENTS ON GRADIENTS OF RESPONSE[1]

MICHAEL A. WALLACH

Harvard University

WE BEGIN with the observation that a person is constantly classifying the contents of the world about him. The child, for instance, is taught what kinds of stimulation may be called green and what ones blue. Later he may likewise learn what modern Italian paintings are futurist in style rather than surrealist, or what people should be called communists. There seem to be certain common features involved in all such classifying, and they may be put as follows. Stimulation is received which the individual arrays along one or more innate or learned dimensions. Such input, along wider or narrower extents of such dimensions, is associated by repeated contiguity with a class name. It is in such a manner as this that the individual comes to learn what parts of the world of stimulation it is appropriate to call of yellow color, politically anarchistic, or impressionistic in style of painting.

But the learning situation is often rife with elements of ambiguity. Was that color blue or green? Its hue was such that one cannot be sure. Here is a painting whose style seems to reflect both futurist and surrealist components. Are this man's activities really of a sort that justifies calling him a communist? With our world of fleeting stimuli, of quick looks, and conflicting opinions as to what class labels are appropriate in what situations, such ambiguities as these would seem to be common indeed. If one is to study how people learn to classify, then, it would seem necessary from the outset to recognize that the world more often than not provides a person with stimulation that does not clearly fit the definition of any one class or label, but rather stands at the boundary of a particular label's applicability; it can be labeled with equal or near equal appropriateness by more than one term. Thus we come to the central question of our inquiry: Do any factors arising from the nature of classification itself influence how it proceeds in an ambiguous situation—i.e., a situation where alternative class names for the given event have been learned to equal degree?

There are at least two mechanisms arising from the way classes are defined which influence the tendency to assign information to a particular class: (*a*) whether a class is defined by inclusion or exclusion and (*b*) whether a class defined by inclusion is contrasted with a class defined by exclusion of the former class or with one also defined by inclusion. Consider the first question. There seem to be two main ways in which one can specify the contents of a class, and our hypothesis is that their psychological implications are quite different. On the one hand, the contents of a class may be specified by inclusion; such-and-such attribute values must be present for the datum to be given a particular class label. If, for instance, something has wings, two legs, and flies through the air, we may appropriately

[1] This monograph is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Harvard University. The author is greatly indebted to Jerome S. Bruner, his principal adviser, for encouragement and aid in this research. He also wishes to thank Gordon W. Allport, John B. Carroll, George Mandler, and Evon Z. Vogt, the other members of his Thesis Committee, and Albert J. Caron, Paul Friedrich, and Frederick Mosteller, all of whom rendered assistance in various ways. Special thanks are due Rainer Weiss for design of the electronic apparatus. This research was conducted while the author held a doctoral research fellowship from the Samuel S. Fels Foundation. Support was also provided by Grant M-1324 from the United States Public Health Service, under the general direction of Jerome S. Bruner.

classify it as a bird. On the other hand, the contents of a class may be specified by exclusion; if such-and-such attribute values are not present, the datum can be called a nonexemplar of the class or classes appropriate to those attribute values. If, for example, an animal does not spend all his time on land and/or in the water, does not have more than two legs, does not lack wings, then we may appropriately classify it as an animal residing somewhere other than on land or sea—and we will be referring here to the same domain of instances that we labeled before as "birds."

Given an inclusion and an exclusion class that both refer to the same domain of instances, our hypothesis is that the range of acceptance limits for placing data in a class will be greater for the former than for the latter. An inclusion class, we suggest, carries with it the implication that it is well represented in the environment—that there are many events for which the class label in question has clear applicability. An exclusion class, on the other hand, is a more vague entity. It is as if people prefer examples of what something is, more than examples of what something is not. This in turn might be expected on the ground, for instance, that it is usually difficult to know how to deal with an event placed in an exclusion class since the event could be any of a variety of things, or that work with an event defined by exclusion usually requires the added cognitive effort of transformation to a definition by inclusion.

Our second suggested classification determinant of range of acceptance limits is whether an inclusion-defined class of interest is being contrasted with a class also defined by inclusion or with the contrast class defined by exclusion of that class of interest. We would propose that reference to some contrast class is always present, implicitly if not explicitly, when we classify, and this for the following reason. In order to avoid being overwhelmed by the confusing richness of attributes that the environment offers the senses, a person must define a class for himself in terms of a limited selection of properties. Consideration of a contrast class, in turn, usually provides one mechanism for such selection. If we make setters the contrast when defining the class "cocker spaniels," for instance, then we may tend to take size as criterial for our definition of the latter class. If, however, we take "all animals other than cockers" as the implied contrast, then we may well define cockers differently. Thus a contrast class too may be defined by inclusion or exclusion, and we hypothesize that this difference is a second determinant of the range of acceptance limits for placing events in an inclusion class of interest. If the contrast class is defined by exclusion of the class of interest, we predict that the acceptance limits for placing events in that inclusion class of interest will be broader than if the contrast class is another inclusion class. This should follow because we expect a contrast class defined by exclusion of a class to be relied on to a lesser extent than an inclusion-defined contrast class, hence giving the class of interest more opportunity in the former case to capture events away from the contrast class.

Variables and Hypotheses Under Study

Dependent Variable

Range of acceptance limits for placing events in a class. The range of acceptance limits for placing events in a class is said to be greater, the greater the number of discriminably different events assigned to that class, and the higher the relative frequency of assignment of each such event to one class rather than to another class or other classes (or the greater the magnitude of some response indicating assignment to a class).

Independent Variables

Definition of a class of interest by inclusion or exclusion (positive or negative class). A domain of data is said to be defined by inclusion, or assigned to a positive class, if the definition of the class specifies the particular attribute values whose presence is definitive of instances of that class. A domain of data is said to be defined by exclusion, or assigned to a negative class, if the definition of the class specifies simply

that the attribute values definitive of instances of some other class or classes are absent—hence, that instances of some positive class or classes are absent.

Presence of a positive or complementary negative contrast class for a positive class of interest. The class considered as a contrast to a positive class of interest may be the negative class complementary to that positive class of interest (if the positive class of interest is called "X," the complementary negative contrast class is "$\bar{X}$"), or the contrast class may be some other positively defined class.

Hypothesized Relationships Between Independent Variables and the Dependent Variable

Hypothesis 1: Concerning the influence of positive vs. negative classes on the range of acceptance limits. (*a*) For a class whose definition is never more than ambiguous, a positive class has a greater range of acceptance limits for placing events in it than does a negative class associated with the same domain of events. (*b*) The same is true for a class whose definition is at least in part certain.

Hypothesis 2: Concerning the influence of a contrast class, and how it is defined, on the range of acceptance limits for the class of interest. (*a*) For a positive class whose definition is never more than ambiguous, the range of acceptance limits for placing events in that class will be greater when contrasted with its complementary negative class than when contrasted with another positive class. (*b*) The same is true for a positive class of interest whose definition is at least in part certain.

It should be noted that we expect these relationships only to hold with regard to data that are generally ambiguous. Data are maximally ambiguous if their relative frequencies of assignment to alternative classes are the same. Data are maximally certain, on the other hand, if their relative frequency of assignment to one class is 100%. We are interested in the determinants of *S*'s classification of generally ambiguous data because such ambiguity is the case in those life situations for which the present hypotheses are supposed to be paradigmatic.

Now a class may be ambiguously associated with some data and associated more certainly with other data, or it may be only ambiguously associated with data. The question thus arises of whether the effects we postulate will hold only with regard to a class which is itself never more than ambiguous, or rather will also hold regarding a class whose definition is at least in part certain. Hence Parts *a* and *b* of each hypothesis. A class is maximally ambiguous, then, if no datum has more than an equal relative frequency of assignment to that class and to one or more other classes, and the greater the number of discriminable data of this kind. On the other hand, a class becomes increasingly certain, the greater the relative frequency of assignment of data to that class rather than to another class or classes, and the greater the number of discriminable data with a high relative frequency of assignment of this kind.

Method

Design of Conditions

We turn now to the matter of testing the hypotheses presented. The experimental requirements are these: (*a*) We must vary the way a class of interest is defined—positively vs. negatively—meanwhile holding the contrast class constant. The logical impossibility of a class of interest and a contrast class both being negative may be noted.[2] Hence, it follows that the contrast class held constant here must be positive, while the class of interest varies from positive to negative. (*b*) We must also vary the way the contrast class is defined, meanwhile holding the type of definition of the class of interest constant. Since we wish to compare the case where the contrast class

[2] It would be logically impossible to have both class of interest and contrast class negative, because two negative classes cannot apply to a domain when the two classes are mutually exclusive (i.e., cannot both apply to the same event at the same time) and are all-inclusive.

is positively defined with the case where it is defined by exclusion of the class of interest, it follows that the class of interest held constant must be defined positively—again to avoid the logically impossible case of contrast class and class of interest both being negative. Further, we want situations where the classes whose acceptance limits we are determining are never more than ambiguous in definition, and other situations where these classes' definitions are in part certain.

Let us refer to a class defined positively as "X" or "Y," and to one defined negatively as "$\bar{X}$" or "$\bar{Y}$." Three conditions, involving association of these classes with a range of data in different ways, will permit us to test the above hypotheses. These conditions are as follows:

The XY condition. Both classes associated with the domain of data are positively defined. The association of Class X with that domain is in part certain, while the association of Class Y with that domain is never more than ambiguous.

The X$\bar{X}$ condition. A positive class and its complementary negative class are associated with the domain of data. The association of the positive class (X) with that domain is in part certain, while the association of the negative class ($\bar{X}$) with that domain is never more than ambiguous.

The $\bar{Y}$Y condition. Again, a positive class and its complementary negative class are associated with the data domain. This time, however, the association of the positive class (Y) with that domain is never more than ambiguous, while the association of its complementary negative ($\bar{Y}$) is in part certain.

How do these conditions permit us to test our hypotheses? Consider Hypothesis 1*a*. If we compare Conditions XY and X$\bar{X}$, we are holding a contrast class constant and positive (X), and varying the definition of an ambiguous class of interest from positive (Y) to negative ($\bar{X}$)—which is what must be done to test that hypothesis. For Hypothesis 1*b*, if we compare conditions XY and $\bar{Y}$Y, we are holding a contrast class constant and positive (Y), and are varying the definition of a class of interest which is in part certain from positive (X) to negative ($\bar{Y}$)—the requirement for testing that, hypothesis. For Hypothesis 2*a*, we again compare Conditions XY and $\bar{Y}$Y, since here the ambiguous positive class of interest is Y, and we compare a positively defined contrast class (X) and a complementary negative contrast class ($\bar{Y}$). Finally, in the case of Hypothesis 2*b*, we consider Conditions XY and X$\bar{X}$ again, for here the positive and, in part, certain class of interest is X, and we compare a positively defined contrast class (Y) and a complementary negative contrast class ($\bar{X}$).

Note that the groups being compared are the same for Hypotheses 1*a* and 2*b*, and for Hypotheses 1*b* and 2*a*. How, then, can each hypothesis in the pair be tested separately if the same conditions are compared for the two hypotheses? The answer lies in consideration of the measures we use to determine range of acceptance limits. One kind of measure involves *S*'s predicting, for each successively presented stimulus event, whether it will be given one or the other class label, or his refusing to predict. Another kind of measure, after *S* has experienced the whole set of associations between various stimulus events and the two class labels, involves *S*'s reporting, for each successively presented datum, whether it was given one or the other class label more than half the time before, or was given each label about half the time before. Provision of the "refusal to predict" response choice in the first kind of measure, and of the "half and half" response choice in the second kind of measure, permit separate testing of each hypothesis in the pairs just noted, even though both in the pair are tested by comparing the same two conditions. This comes about because these response choices allow the relative frequency of assignment of data to Class X to vary independently of the relative frequency of assignment of data to Class Y for the XY condition; likewise, the relative frequencies of data assignment to Classes X and $\bar{X}$ can vary independently in the X$\bar{X}$ condition, and those to $\bar{Y}$ and Y can vary independ-

ently in the $\bar{Y}Y$ condition. These frequencies are independent since there is always a third choice which *S* can make in any of the groups: he can refuse to offer a prediction or can say each label occurred about half the time before. (Let us abbreviate responses of the latter two sorts as "O" judgments.) Calling the two class labels in each case (i.e., X vs. Y, X vs. $\bar{X}$, $\bar{Y}$ vs. Y) Labels 1 and 2, respectively, we hence see that assignment of input to Label 1 can change in relative frequency without a reciprocal change in relative frequency of assignment of input to Label 2 (or vice versa), because there is a "cushion" in between: namely, "O" judgments.

Consider now what each of the hypotheses predicts concerning these relative frequencies. According to Hypothesis 1*a*, comparing XY and $X\bar{X}$, the relative frequency of assignment of data to Class Y should be greater than that to Class $\bar{X}$. By Hypothesis 1*b*, comparing XY and $\bar{Y}Y$ conditions, the relative frequency of assignment of data to Class X should be greater than that to Class $\bar{Y}$. By Hypothesis 2*a*, again comparing XY and $\bar{Y}Y$ conditions, the relative frequency of assignment of data to Class Y should be greater for the $\bar{Y}Y$ group, and less for the XY group. Finally, by Hypothesis 2*b*, again comparing the XY and $X\bar{X}$ groups, the relative frequency of assignment of data to Class X should be greater in the $X\bar{X}$ group, and less in the XY group. It becomes clear, then, that the four hypotheses can be independently tested if, as is the case, the relative frequencies of data assignment to Classes X vs. Y, X vs. $\bar{X}$, and $\bar{Y}$ vs. Y, are independent.

Experimental Procedure

Light intensity was used as the dimension of stimulus input, since it was a simple one and changes in its values could be easily created and measured. The *S*'s training regimen involved associating brightnesses with labels or class names.

The questions of defining a class positively vs. negatively, and providing a contrast class defined positively or defined by exclusion of the positive class of interest, were handled by varying the kinds of class labels provided. For Condition XY, where both labels were to be positive, when a brightness was to be linked with Label 1, a bulb lit up the letter "X"; and when a brightness was to be linked with Label 2, a different bulb lit up the letter "Y"—the "X" and "Y" bulbs being placed one on each side of the screen where brightnesses were presented. When there are separate "X" and "Y" bulbs and one of them lights each time, we thereby imply to *S* that there are two classes each defined by particular values of the brightness attribute—i.e., two positive classes.

For Conditions $X\bar{X}$ and $\bar{Y}Y$ where, on the other hand, one label was to be positive and the other its complementary negative, when a brightness was to be linked with the positive label, a bulb lit; while when a brightness was to be linked with the negative label, that same bulb failed to light—the single bulb being placed underneath the screen where brightnesses were presented, and no other bulb being present. When there is just one bulb which either lights or fails to light each time, we thereby imply to *S* that there is one class defined by particular brightness values and another class defined simply by absence of those values—i.e., a positive class and a complementary negative class, respectively.

Note that the actual stimulus inputs of brightnesses and the relative frequencies of particular associations of brightnesses with one of two possible class labels, were alike in all conditions. What varied were the instructions provided by the light labels as to how the classes were defined, while the structures of the three pairs of classes were exactly the same.

The number of different brightness values used was 16, these being spaced on a scale of just noticeable differences (see Table 1).[3] The total number of presentations of brightnesses during the training series was 160, each brightness coming up 10 times and the whole sequence being uniquely random, but every brightness occurring twice in each block of 32 presentations. The relative fre-

[3] The JND's were established by trials of the following sort: A rheostat was slowly turned upward from minimal brightness (or downward from maximal brightness), brightness intensity level being recorded every time *S* reported a just noticeable difference. Two upward series and two downward series were given in single alternation to any one *S*. Sixteen was about the maximal number of JND's detectable within the brightness range available from the rheostat. The *E* ran this procedure on himself and also on two college student *S*s. Final brightness intensity settings were arrived at by taking all series in which 16 brightness values had been provided by the judge (i.e., 15 JND's) and finding the median of the successive intensity settings for each of the 16 values. Sixteen brightness values were provided in three of the four series by one judge, and in two of the four series by each of the two other judges. (Fourteen or 15 different values were provided in the remaining series.)

TABLE 1

BRIGHTNESS LEVELS USED

Brightness Intensity (In DC ma. using a 60-w. bulb)	Brightness No.
1.00 (Brightest)	1
.98	2
.96	3
.88	4
.84	5
.80	6
.76	7
.72	8
.68	9
.63	10
.57	11
.53	12
.48	13
.43	14
.40	15
.36 (Darkest)	16

quencies of association of each brightness with each of the two labels used in any one condition are shown in Fig. 1. As indicated on that figure, toward the brighter end of the brightness dimension all presentations were always associated with one label, while toward the darker end, all presentations were equally often associated with each of the two labels. The sequence of occurrence of the two labels in any condition was uniquely random, and was the same for the label of the same number (1 or 2) in all the conditions.

Seventy persons participated in the study, the number of university student *S*s in Groups XY and $X\bar{X}$ being 23 each, while Group $\bar{Y}Y$ contained 24 *S*s. Procedure was the same for all three groups except for the labeling differences among these groups as discussed above. This procedure fell into three parts and was arrived at as a result of pilot investigation.

Training series. On each of the 160 trials, one of the 16 brightnesses was flashed for 0.5 sec. in a uniquely random order. After each brightness was presented, *S* pushed one of three buttons, indicating, if in Group XY, that he bets the "X" light will go on, the "Y" light will go on, or that he refrains from making a bet, respectively; if in Groups $X\bar{X}$ or $\bar{Y}Y$, that he bets the single light will go on ("Yes"), the single light won't go on ("No"), or that he refrains from making a bet, respectively. When *S* pushed any button, the "X" or "Y" light went on for 1.0 sec. if in Group XY, or the single light went on for 1.0 sec. or remained off if in Groups $X\bar{X}$ or $\bar{Y}Y$—which of the two class labels occurred depending on the brightness that occurred on that trial. After a pause of 5.5 sec. following a button push, the next brightness in the sequence flashed, thereby beginning the next trial. Thus, in the course of the 160 training trials, each brightness was linked with one or both labels according to the pattern of association diagrammed in Fig. 1.

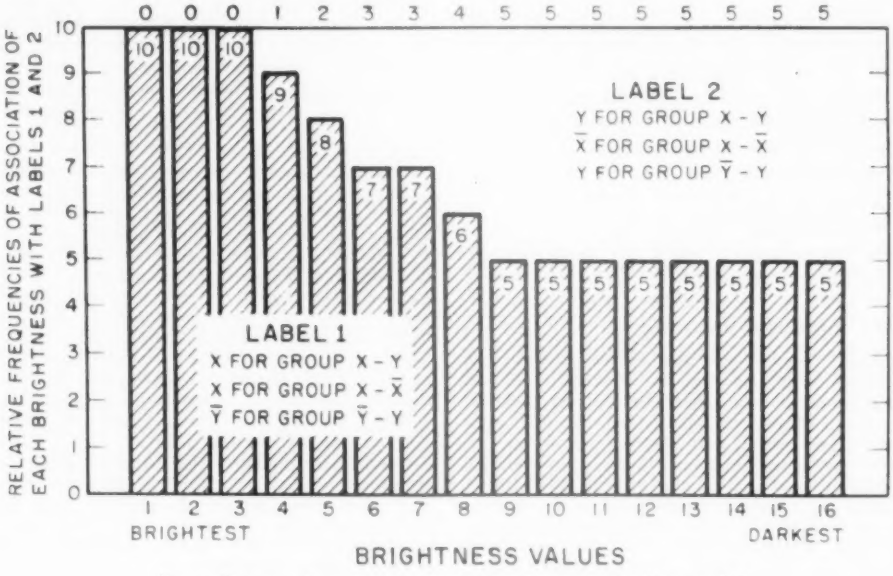
0 0 0 1 2 3 3 4 5 5 5 5 5 5 5 5
10 10 10 9 8 7 7 6 5 5 5 5 5 5 5 5
LABEL 2
Y FOR GROUP X - Y
X̄ FOR GROUP X - X̄
Y FOR GROUP Ȳ - Y
LABEL 1
X FOR GROUP X - Y
X FOR GROUP X - X̄
Ȳ FOR GROUP Ȳ - Y
RELATIVE FREQUENCIES OF ASSOCIATION OF EACH BRIGHTNESS WITH LABELS 1 AND 2
0 1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
BRIGHTEST
DARKEST
BRIGHTNESS VALUES

FIG. 1. Degrees of association of each brightness with Labels 1 and 2.

Test with brightnesses. On each of 32 trials, one of the 16 brightnesses was flashed for 0.5 sec. in a uniquely random order, each brightness occurring twice during the series. After each brightness was presented, *S*, if in Group XY, indicated, by pushing a button labeled "More," that he thinks this brightness turned out to be X more than half of the times it was previously presented or turned out to be Y more than half those times, and then told *E* verbally whether he meant X or Y more than half the time before; or *S* might push another button labeled "Equal," to indicate that he thinks this brightness turned out to be X about half those times and Y about half those times. If in Groups $X\bar{X}$ or $\bar{Y}Y$, *S* indicated, by pushing the button labeled "More," that he thinks this brightness caused the light to go on more than half of the times it was previously presented or caused the light to stay off more than half of those times, and then told *E* verbally which he meant; or *S* might push the other button labeled "Equal" to indicate that he feels this brightness caused the light to go on about half those times and to stay off the other half. No label lights went on during this test series. Five and one-half seconds after *S* pushed a button, the next brightness in the sequence flashed, thus starting the next trial.[4]

Test without brightnesses. The *S* was shown a sheet of paper along which a horizontal line was drawn to represent the dimension of brightnesses to which he had been exposed, the word "Darkest" appearing at one end of the line and "Brightest" at the other.[5] Along a vertical line at the left, divided into 10% steps ranging from 0% to 100%, appeared the legend "Proportion of times you think each brightness was labeled 'X'" if for Group XY; while the legend for Groups $X\bar{X}$ and $\bar{Y}Y$ read "Proportion of times you think each brightness made the light go on." The *S* was requested to draw a gradient indicating the proportion of times he thinks each brightness was labeled "X" (Group XY), or made the light go on (Groups $X\bar{X}$ and $\bar{Y}Y$), for all parts of the brightness dimension.

The choice measure in the training series indicated the learning of labeling tendencies. The choice measure in the test with brightnesses provided a more perceptually based indication of the labeling tendencies thus learned (since the brightnesses were presented again), while the graph measure in the test without brightnesses provided a more conceptually based index of the same thing (since the brightnesses were not presented). Latency measures in the training series and the test with brightnesses told us how confident *S* was in his labeling decisions.

Apparatus

The time lengths for presentation of the brightnesses and of the label lights, and for the interval between *S*'s pushing of a button and the occurrence of the next brightness in the sequence, were automatically controlled by timers. Pre-established programs on stepping switches automatically controlled the sequences in which the brightnesses occurred and in which the two labels in any condition occurred. The brightness differences were differences in the voltage settings of 16 rheostats, some one of which controlled the lighting of a common incandescent bulb for each trial. This high degree of automatic control of the sequence and timing of stimulus and label materials was required in order to free *E* to record the various items of data that *S* provided, while still maintaining strict temporal control over presentation of stimulus input and label materials.

The sequence of events in time is shown in Fig. 2, and the block diagram of Fig. 3 indicates the various apparatus components and how they were linked so as to result in this sequence. The push-buttons for the test series were mounted on one portable board while those for the training series were mounted on another, and one set of push-buttons was placed out of *S*'s sight when he was

[4] Use of the two buttons noted in this test without brightnesses, rather than providing three buttons for each of the three alternative decisions, was instituted as an added control to insure that *S* was pushed toward veridicality of report rather than away from it. A single button labeled "More" and another one labeled "Equal" imply to *S* that there might well be some brightnesses linked to one label more than half the time, and about as many other brightnesses linked to each label equal proportions of the time—which is in fact the case. The implication tendered, in other words, is that there is warrant to press the "Equal" button on about half of the test trials. To follow this implication would lead *S* toward veridicality. If nonveridicality is obtained, then, in that Label 2 is thought to apply more often than was the case, this tendency was not fostered by the mere physical conditions of the test itself.

The instructions given the various groups again insured that *S* was pushed toward veridicality of report rather than away from it, in that the examples given of kinds of label-input associations told only of Label 1 being linked with a brightness shade all the times that shade is presented, or most of the times it is presented, or about half the times. No example was mentioned where Label 2 was said to be linked with a brightness shade more than half the time, although of course this possibility was not ruled out. Once again, if nonveridicality is obtained in that Label 2 is thought to apply more than half the time to some brightness, then this decision was not fostered by bias in the communicated instructions.

[5] For half the *S*s in each group, "Darkest" appeared at the left and "Brightest" at the right end of the line, while for the other half these word positions were reversed.

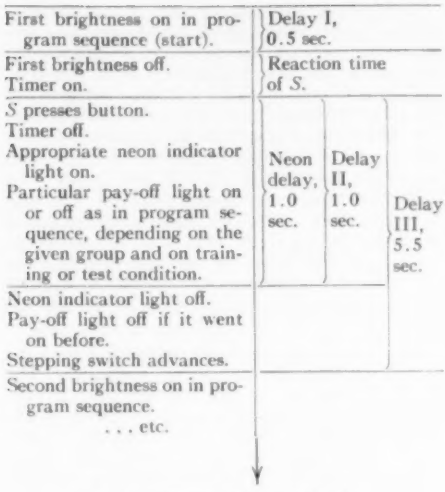
First brightness on in program sequence (start).
Delay I, 0.5 sec.
First brightness off.
Timer on.
Reaction time of S.
S presses button.
Timer off.
Appropriate neon indicator light on.
Particular pay-off light on or off as in program sequence, depending on the given group and on training or test condition.
Neon delay, 1.0 sec.
Delay II, 1.0 sec.
Delay III, 5.5 sec.
Neon indicator light off.
Pay-off light off if it went on before.
Stepping switch advances.
Second brightness on in program sequence.
. . . etc.

FIG. 2. Sequence of events in time.

using the other. Neon indicator lights were mounted on *E*'s panel to tell him which button *S* pushed on that trial. The neon delays resulted in a neon light's staying on for 1 sec. after *S* released the button, so *E* was sure to see which light went on. The *E* began the brightness sequence in the training or test series by pressing the start button, and from then on the series proceeded automatically. When the end of the series was reached and *E* wished to set up the sequence for the next part of the experiment or for the next *S*, *E* returned the program sequences to their starting position by pressing the reset button, which homes all the stepping switches. The timer utilized was a Standard Electric Time Co. product, Type S-1. A regular 60-w. lamp of white color was the brightness intensity bulb, and it was mounted behind a circle of translucent blue celluloid 1.5 in. in diameter, so brightness differences were various intensities of blue light. Pay-off lights were 3-w. lamps mounted in pilot light fixtures.

The experiment was run in semidarkness, room lights being off and the only source of illumination other than the brightness intensities and pay-off lights being a small 25-w. reading lamp placed well behind a black curtain on *E*'s desk.[6] The *S* sat 3 ft. from the light panel and at eye level with it. Figure 4 shows a sketch of the apparatus from *S*'s point of view.

RESULTS

Tables 2–6 summarize the results obtained using the measures outlined above. It will be noticed that when results are presented concerning various parts of the whole

[6] This semidarkness was maintained during the training series, during the instructions for the test with brightnesses, and during this latter test itself.

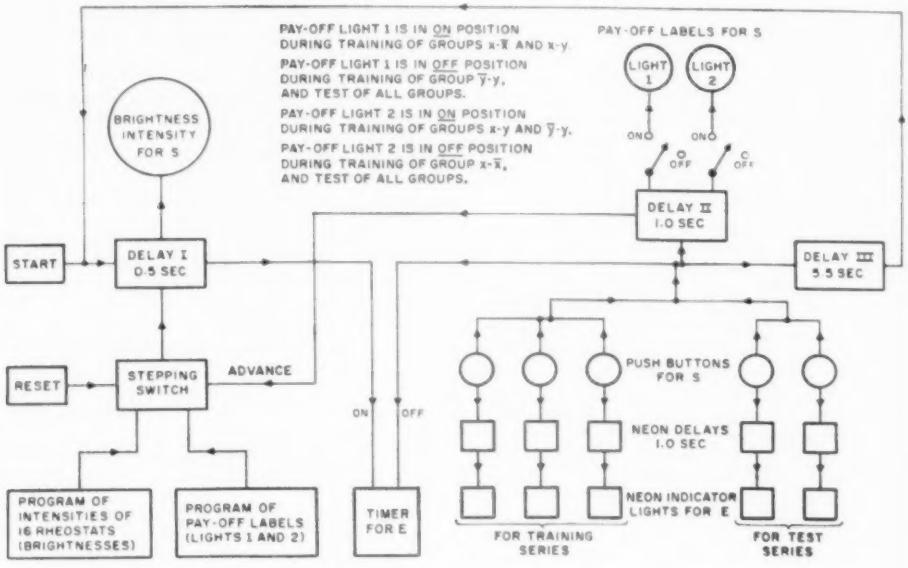
PAY-OFF LIGHT 1 IS IN ON POSITION DURING TRAINING OF GROUPS x-x̄ AND x-y.
PAY-OFF LIGHT 1 IS IN OFF POSITION DURING TRAINING OF GROUP ȳ-y, AND TEST OF ALL GROUPS.
PAY-OFF LIGHT 2 IS IN ON POSITION DURING TRAINING OF GROUPS x-y AND ȳ-y.
PAY-OFF LIGHT 2 IS IN OFF POSITION DURING TRAINING OF GROUP x-x̄, AND TEST OF ALL GROUPS.
PAY-OFF LABELS FOR S
LIGHT 1
LIGHT 2
ON
OFF
BRIGHTNESS INTENSITY FOR S
DELAY II 1.0 SEC
DELAY III 5.5 SEC
START
DELAY I 0.5 SEC
RESET
STEPPING SWITCH
ADVANCE
PUSH BUTTONS FOR S
NEON DELAYS 1.0 SEC
NEON INDICATOR LIGHTS FOR E
PROGRAM OF INTENSITIES OF 16 RHEOSTATS (BRIGHTNESSES)
PROGRAM OF PAY-OFF LABELS (LIGHTS 1 AND 2)
TIMER FOR E
FOR TRAINING SERIES
FOR TEST SERIES

FIG. 3. Block diagram of the components of the apparatus.

dimension of 16 brightnesses, we fractionate that dimension into four parts of four brightnesses each and perform calculations for each part separately, as well as examining the effect for all 16 brightnesses taken as a whole. Our reason for breaking the brightness dimension down into no more and no less than four parts, when examining trends across the dimension, is that four was the modal and median number of different brightnesses which our *S*s discriminated in the course of brightness presentations in the two random series. After finishing the test with brightnesses and before starting the test without brightnesses, each *S* was asked how many different brightnesses he thought had been used—and the mode and median reply they gave was four. It hence seemed most psychologically meaningful to take this estimate as the number of segments into which to break the brightness dimension for analysis.

Turning now to the findings in Tables 2–5: For any measure, the results for the three groups of the experiment were first treated by the Kruskal-Wallis ranked analysis of variance (Siegel, 1956). If the probability of the distributions being from a common population was more than .10, the finding was considered not significant. If this probability was more than .05 but less than or equal to .10, the probability level was entered since it might be suggestive; but since the standard convention for statistical significance was not met, no further test could be made for the measure in question. Finally, if the probability level was less than or equal to .05, we proceeded to determine whether this over-all significance of the ranked analysis of variance was due to the difference in the distributions of Groups $X\bar{X}$ vs. XY, XY vs. $\bar{Y}Y$, or both.[7] For comparisons of two groups, the Mann-Whitney *U* test (Siegel, 1956) was employed. When the measure involves Label 1 or Label 2, our hypothesis specifically pre-

[7] Such a procedure is analogous to following up the finding that a parametric analysis of variance is significant by using Tukey's test for a significant gap to determine which intergroup differences cause the significant over-all effect. One is, of course, warranted to do tests concerning a pair of the groups only if the over-all effect for the three groups is significant.

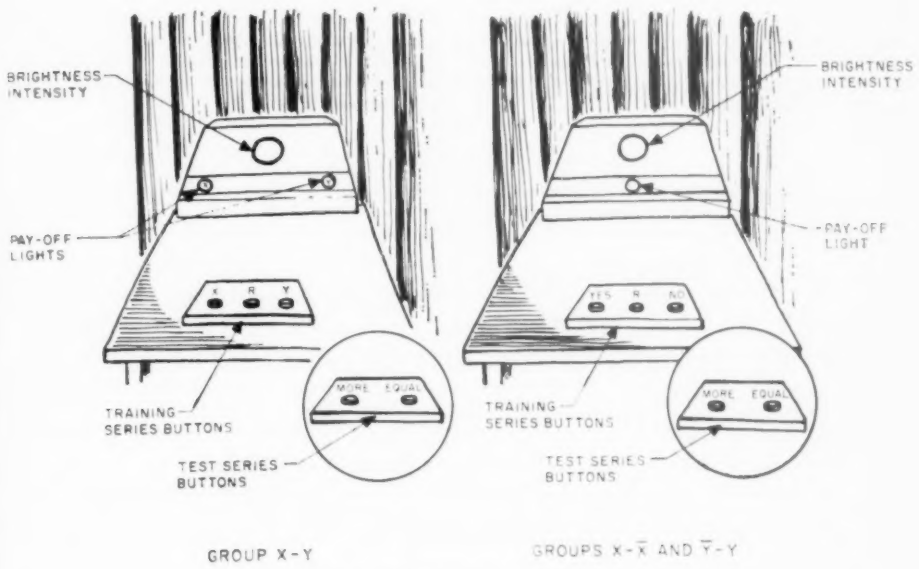

FIG. 4. The apparatus as seen by *S*.

dicts the direction of the difference between Groups $X\bar{X}$ and XY, or between Groups XY and $\bar{Y}Y$. Hence, when a Mann-Whitney test concerns Label 1 or Label 2, a one-tail test was made. When, on the other hand, the measure concerns refraining, "equal" judgments, or the indifference point, we have alternative hypotheses which could predict either direction of a difference, and hence a two-tail test was made.[8]

We offer the reminder that our hypotheses are expected to hold most strongly concerning the more ambiguous half of the brightness dimension—namely, Brightnesses 9–16—since this is the area at the boundary of the category defined by Label 1. Our findings indicate that such, by and large, is the case: The hypotheses noted earlier are usually confirmed, in whole or in part, for Segments 9–12 and 13–16 of the brightness dimension, and are usually not confirmed for Portions 1–4 and 5–8. It hence requires a quite strong effect over Brightnesses 9–16 for our hypotheses to be confirmed when the whole dimension—Brightnesses 1–16—is considered as a unit: yet such confirmation often occurs to at least some extent.[9]

Label Learning (Choice Measure, Training Series)

Tests of the frequencies of choice of Label 1, Label 2, and of the refraining button during the last 32 trials of the training series, are presented in Table 2. From the choice frequencies for Label 1, we note that Hypothesis 1*b*, predicting $XY > \bar{Y}Y$, is supported at the .07 level for Brightnesses 13–16, and Hypothesis 2*b*, predicting $X\bar{X} > XY$, is supported at the .01 level for those brightnesses. Neither hypothesis is supported for other parts of the brightness dimension. The choice frequencies for Label 2 indicate that Hypothesis 1*a*, predicting $X\bar{X} < XY$, is supported at the .09 level for Brightnesses 9–12, and at the .001 level for Brightnesses 13–16. Hypothesis 2*a*, predicting $XY < \bar{Y}Y$, is supported at the .06 level for 9–12, but not for 13–16. The over-all effect for Brightnesses 1–16 supports Hypothesis 1*a* at the .01 level, and

[8] Whitney's extension of the *U* test to three groups (Mosteller & Bush, 1954) might seem appropriate for Tables 2–5, since this test permits one to evaluate the hypothesis that the distributions of three groups fall in a particular order. It could not be used, however, because there are large numbers of ties for scores among our three groups on a measure, and application of this technique when there are large numbers of intergroup ties involves theoretical difficulties.

[9] The measure of central tendency used is the median. Occasionally, however, significant effects occur which are not reflected in the medians for the groups in question, and in such cases means have been added.

TABLE 2

MEDIAN FREQUENCY OF CHOICE OF LABELS 1 AND 2 AND THE REFRAIN BUTTON (R) DURING THE LAST 32 TRIALS OF THE TRAINING SERIES

Brightness	Group $X\bar{X}$ (N = 23)			Group XY (N = 23)			Group $\bar{Y}Y$ (N = 24)			Kruskal-Wallis Analysis of Variance by Ranks			Mann-Whitney *U* Tests					
													Between Groups $X\bar{X}$ and XY			Between Groups XY and $\bar{Y}Y$		
	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R
1–4, median	8.0	0.0	0.0	7.0	0.0	0.0	6.5	0.0	0.0	N.S.	N.S.	N.S.						
5–8, median	5.0	0.0	1.0	7.0	1.0	0.0	5.0	1.0	0.0	N.S.	N.S.	N.S.						
9–12, median	4.0	1.0	2.0	5.0	2.0	1.0	3.5	3.0	0.0	N.S.	$\leq .05$	N.S.		$\leq .09$			$\leq .06$	
13–16, median	3.0	3.1	2.0	1.0	6.0	1.0	0.0	6.0	0.0	$\leq .01$	$\leq .001$	$\leq .02$	$\leq .01$	$\leq .001$	$\leq .03$	$\leq .07$	N.S.	N.S.
1–16, median	19.0	5.0	7.0	20.0	9.0	2.0	15.5	11.0	2.0	$\leq .08$	$\leq .001$	$\leq .08$		$\leq .01$			$\leq .03$	
1–16, mean	19.7			18.9			16.2											

supports Hypothesis 2*a* at the .03 level. Regarding the frequencies of refraining, we note that $X\bar{X} > XY$ at the .03 level for Brightnesses 13–16, a finding that supports Hypothesis 1*a* and infirms Hypothesis 2*b*. The findings of no significant difference for comparisons of other groups over Brightnesses 9–12 and 13–16 may be supportive, in general, of all hypotheses in question, or may, on the other hand, simply be due to nonconfirmation of these hypotheses. Interpretation of failure to find a significant difference between two groups in frequencies of refraining must always remain ambiguous in this way.

Perceptually Based Indication of Labeling Tendencies (Test With Brightnesses)

Table 3 concerns the frequencies of choice of Label 1, Label 2, and of the "Equal" button during the test with brightnesses. With regard to choice of Label 1, we find Hypothesis 2*b*, that $X\bar{X} > XY$, is supported at the .001 level for Brightnesses 13–16, but Hypothesis 1*b* is not supported for those brightnesses, and neither hypothesis is supported for other brightnesses. For choice of Label 2, we find that Hypothesis 1*a*, predicting that $X\bar{X} < XY$, is supported at the .001 level for Brightnesses 9–12 and 13–16, while Hypothesis 2*a*, pre-

TABLE 3

MEDIAN FREQUENCY OF CHOICE OF LABELS 1 AND 2 AND THE REFRAIN BUTTON (R) DURING THE TEST WITH BRIGHTNESSES

Brightness	Group $X\bar{X}$ (N = 23)			Group XY (N = 23)			Group $\bar{Y}Y$ (N = 24)		
	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R
1–4, median	8.0	0.0	0.0	8.0	0.0	0.0	7.5	0.0	0.0
5–8, median	6.0	0.0	2.0	6.0	0.0	2.0	4.5	1.0	2.0
5–8, mean		0.1			0.7			1.0	
9–12, median	2.0	0.0	5.0	2.0	3.0	3.0	1.0	4.0	2.5
13–16, median	1.0	0.0	6.0	0.0	7.0	1.0	0.0	8.0	0.0
1–16, median	16.0	0.0	15.0	16.0	11.0	7.0	12.5	12.0	7.0
1–16, mean	17.1			14.4			13.2		

Brightness	Kruskal-Wallis Analysis of Variance by Ranks			Mann-Whitney U Tests: Between Groups $X\bar{X}$ and XY			Mann-Whitney U Tests: Between Groups XY and $\bar{Y}Y$		
	Label 1	Label 2	R	Label 1	Label 2	R	Label 1	Label 2	R
1–4	N.S.	N.S.	N.S.						
5–8	N.S.	≤.01	N.S.		≤.01			N.S.	
9–12	N.S.	≤.001	≤.001		≤.001	≤.001		≤.01	N.S.
13–16	≤.001	≤.001	≤.001	≤.001	≤.001	≤.001	N.S.	≤.02	≤.06
1–16	≤.02	≤.001	≤.001	≤.09	≤.001	≤.001	≤.09	≤.01	N.S.

dicting that $XY < \bar{Y}Y$, is supported at the .01 level for 9–12, and at the .02 level for 13–16. Hypothesis 1*a* is also supported for Brightnesses 5–8 at the .01 level, and both hypotheses are supported for the brightness dimension as a whole—Hypothesis 1*a* at the .001 level, Hypothesis 2*a* at the .01 level. Concerning "equal" judgments, Hypothesis 1*a*, that $X\bar{X} > XY$, is supported at the .001 level for both 9–12 and 13–16, and this same finding infirms Hypothesis 2*b*. Also, $XY > \bar{Y}Y$ at the .06 level for Brightnesses 13–16, a result that supports Hypothesis 2*a* and infirms Hypothesis 1*b*. The lack of significant difference between XY and $\bar{Y}Y$ for 9–12 may support both Hypotheses 1*b* and 2*a*, or simply reflect their nonconfirmation in this instance. Hypothesis 1*a* is also supported by the finding that $X\bar{X} > XY$ for the whole brightness dimension, at the .001 level.

Conceptually Based Indication of Labeling Tendencies (Test Without Brightnesses)

It is of interest to compare the individual's judgments of the past relationship between *perceived* brightnesses and class labels—the test just described above—with his judgments of the past relationship between points along the brightness dimension as he *conceptualizes* it, and those labels. The test without brightnesses provides an index of the latter, and our results indicate a high degree of match between an *S*'s judgments on the two tests—and hence between his implicit (more perceptual) and

TABLE 4

MEDIAN AMOUNT OF GRAPH AREA ASSIGNED TO LABELS 1 AND 2, AND TO 50% JUDGMENTS DURING THE TEST WITHOUT BRIGHTNESSES[a]

Brightness	Group $X\bar{X}$ (N = 23)			Group XY (N = 23)			Group $\bar{Y}Y$ (N = 24)		
	Label 1	Label 2	50%	Label 1	Label 2	50%	Label 1	Label 2	50%
1–4	350.0	50.0	0.0	370.0	30.0	0.0	344.5	55.5	0.0
5–8	273.0	127.0	0.0	262.0	138.0	0.0	230.0	170.0	0.0
9–12	224.0	176.0	0.0	172.0	228.0	0.0	145.0	255.0	0.0
13–16	217.0	183.0	0.0	60.0	340.0	0.0	55.0	345.0	0.0
13–16, mean			54.3			8.7			6.3
1–16	1096.0	504.0	100.0	867.0	733.0	50.0	756.5	843.5	50.0

Brightness	Kruskal-Wallis Analysis of Variance by Ranks			Mann-Whitney U Tests: Between Groups $X\bar{X}$ and XY			Mann-Whitney U Tests: Between Groups XY and $\bar{Y}Y$		
	Label 1	Label 2	50%	Label 1	Label 2	50%	Label 1	Label 2	50%
1–4	N.S.	$\leq .05$	N.S.		N.S.			$\leq .01$	
5–8	$\leq .03$	$\leq .03$	N.S.	N.S.	N.S.		$\leq .01$	$\leq .01$	
9–12	$\leq .001$	$\leq .001$	N.S.	$\leq .001$	$\leq .001$		$\leq .02$	$\leq .02$	
13–16	$\leq .001$	$\leq .001$	$\leq .01$	$\leq .001$	$\leq .001$	$\leq .01$	N.S.	N.S.	N.S.
1–16	$\leq .001$	$\leq .001$	$\leq .07$	$\leq .001$	$\leq .001$		$\leq .001$	$\leq .001$	

[a] 1 = 1% of a single brightness' relative frequency of occurrence.

explicit (more conceptual) models as to what went on during the training series. Table 4 presents the findings concerning the test without brightnesses: For Label 1, we find that $X\bar{X} > XY$ at the .001 level for Brightnesses 9–12 and 13–16, and also at the .001 level for the brightness dimension as a whole. Since results for Labels 1 and 2 are interdependent in the test without brightnesses, the above finding is also reflected, for Label 2, in the result that $X\bar{X} < XY$ at the .001 level for 9–12, 13–16, and the whole brightness dimension. These findings may support Hypotheses 1*a*, 2*b*, or both. Analogously, that $XY > \bar{Y}Y$ at the .02 level for Label 1, Brightnesses 9–12, is reflected also in $XY < \bar{Y}Y$ at the .02 level for Label 2, over the same brightness range—results which may support Hypotheses 1*b*, 2*a*, or both. (This effect is also present at the .001 level over the whole brightness range.) Amounts of graph area for XY and $\bar{Y}Y$ are not different, however, for Label 1 or 2 over Brightnesses 13–16, which is contrary to Hypotheses 1*b* and 2*a*. Regarding the amount of graph area assigned to "equal" judgments, that $X\bar{X} > XY$ at the .01 level supports Hypothesis 1*a* and infirms Hypothesis 2*b*. The absence of other significant differences for 9–12 and 13–16 may support all the hypotheses in question, or simply be due to their nonconfirmation in these instances.

Indifference Points

Concerning the relative positions of the median indifference points, that $X\bar{X} > XY$ at the .01 level for the test with brightnesses and for the test without brightnesses (Table 5) confirms Hypothesis 1*a* and infirms Hypothesis 2*b* in both cases; and that $XY > \bar{Y}Y$ at the .01 level for the test without brightnesses confirms Hypothesis 2*a* and infirms Hypothesis 1*b*. The nonsignificant differences between XY and $\bar{Y}Y$ for the test with brightnesses, and between $X\bar{X}$ and XY, and XY and $\bar{Y}Y$ for the training series, again may support all the relevant hypotheses or be an effect of their nonconfirmation.

Reaction Times, Training Series, and Test Without Brightnesses

If the reaction time measure is a valid index of judgment confidence, then latencies for predictions of Label 1 should drop as

TABLE 5

MEDIAN INDIFFERENCE POINT ON THE BRIGHTNESS DIMENSION[a]

Part of Experiment	Groups			Kruskal-Wallis Analysis of Variance by Ranks	Mann-Whitney *U* Tests	
	$X\bar{X}$	XY	$\bar{Y}Y$		Between Groups $X\bar{X}$ and XY	Between Groups XY and $\bar{Y}Y$
Last 32 trials of training series, median	10.0 ($N = 21$)	11.5 ($N = 17$)	9.0 ($N = 14$)	$\leq .09$		
Test with brightnesses, median	11.0 ($N = 23$)	8.5 ($N = 21$)	8.5 ($N = 21$)	$\leq .001$	$\leq .01$	N.S.
Test without brightnesses, median	11.3	8.5	8.5	$\leq .01$	$\leq .01$	$\leq .01$
Test without brightnesses, mean	11.5 ($N = 22$)	9.0 ($N = 23$)	8.0 ($N = 24$)			

[a] Each individual score is the median brightness assigned to the 50% judgment category by each S.

one moves from Brightness 8 to Brightness 1, since the association of brightness with Label 1 becomes more and more probable for all groups.[10] Further, considering predictions of Label 2 for Brightnesses 9–16, Hypothesis 1*a* predicts a sharper drop in latency (i.e., a sharper degree of slope), as one moves from 9 to 16, for Group XY than for Group $X\bar{X}$; and Hypothesis 2*a* predicts a sharper drop in latency over this brightness range for Group $\bar{Y}Y$ than for Group XY. We wished to measure, then, the degree of slope in Label 1 reaction times from Brightnesses 5–8 to Brightnesses 1–4, and in Label 2 reaction times from Brightnesses 9–12 to 13–16. As the subject score for use in analysis of these drops in latency, a measure was needed which would be free of individual differences in absolute level of reaction time, since the absolute difference between Label 1 latencies for 5–8 and 1–4, or between Label 2 latencies for 9–12 and 13–16, would be less for *S*s whose reaction times are shorter in general. This would come about because of a "floor effect": There will be less variability in Label 1 latencies for 1–4 than for 5–8, and less variability in Label 2 latencies for 13–16 than for 9–12. As a result, faster Label 1 responders for 5–8 will show less of a slope to 1–4 than the slower responders simply because the fast people's Label 1 responses to 1–4 cannot become much quicker than their responses to 5–8, while those of the slow people can. The same holds with regard to Label 2 responders for 9–12 and 13–16.

To correct for this floor effect, it hence was necessary to make the slope score a relative one—a ratio of the amount of Label 1 latency drop from 5–8 to 1–4 divided by the total possible amount of drop; and similarly, a ratio of the amount of Label 2 latency drop from 9–12 to 13–16 divided by the total possible amount of drop.

[10] Slope predictions for Label 1 from Brightnesses 9–16 are difficult to make since the greatest concentration of reaction times for Label 1 is expected to be from Brightnesses 1–8, approximately. Such predictions would concern Hypotheses 1*b* and 2*a*.

With regard to the statistical tests used for reaction time data: The first matter of interest was to determine whether the degree of downward slope in reaction time found for each group—slope for Label 1 responses from 5–8 to 1–4, or slope for Label 2 responses from 9–12 to 13–16—was significant or not. In the case of Label 1 responses, our prediction is that there will be a significant downward slope in reaction time from Brightnesses 5–8 to Brightnesses 1–4, for all groups. With regard to Label 2 responses, we expect that there will not be a significant downward slope from 9–12 to 13–16 for Group $X\bar{X}$, but that there will be a moderately significant downward slope for Group XY, and an even more significant slope of this kind for Group $\bar{Y}Y$. In the cases where we predict a difference between latencies for one group of brightnesses and another, then, our hypothesis predicts that this slope will take a particular direction: downward from 5–8 to 1–4 for Label 1 responses, and downward from 9–12 to 13–16 for Label 2 responses. Since we are thus asking whether the slope is significantly different from zero in a certain direction, a one-tail test of the latency drops is in order. Wilcoxon matched-pairs signed-ranks tests (Siegel, 1956) were used for this purpose, and their results are entered in Table 6.

Concerning the drop in latency for Label 1 responses from 5–8 to 1–4, we find the degree of this slope to be significant for all three groups during the last 32 trials of the training series, and also significant for all groups during the test with brightnesses. The significance levels met for various groups here range from .05 to .01. We also have no special reason to expect the degrees of slope to differ among the three groups; for all conditions, confidence should be greater in applying Label 1 to Brightnesses 1–4 than to Brightnesses 5–8, since association of brightnesses with Label 1 is far from certain for 5–8, and almost certain for 1–4. To determine whether there were in fact no slope differences between groups, a Kruskal-Wallis analysis of variance by ranks was

TABLE 6

MEDIAN SLOPE OF REACTION TIME FOR LABEL 1 RESPONSES FROM MIDDLE OF BRIGHTNESS DIMENSION TO BRIGHTEST EXTREME, AND FOR LABEL 2 RESPONSES FROM MIDDLE OF BRIGHTNESS DIMENSION TO DARKEST EXTREME[a]

Part of Experiment	Group $X\bar{X}$		Group XY		Group $\bar{Y}Y$	
	Label 1	Label 2	Label 1	Label 2	Label 1	Label 2
Last 32 trials of training series	.07 ($N = 23$)	.03 ($N = 15$)	.08 ($N = 23$)	.09 ($N = 22$)	.07 ($N = 24$)	.19 ($N = 23$)
Test with brightnesses	.16 ($N = 23$)	.25 ($N = 4$)	.27 ($N = 22$)	.20 ($N = 20$)	.29 ($N = 23$)	.40 ($N = 24$)

Part of Experiment	Wilcoxon Matched-Pairs Signed-Ranks Tests						Kruskal-Wallis Analysis of Variance by Ranks	
	For Group XX		For Group XY		For Group $\bar{Y}Y$			
	Label 1	Label 2	Label 1	Label 2	Label 1	Label 2	Label 1	Label 2
Last 32 trials of training series	$\leq .03$	N.S.	$\leq .025$	$\leq .10$	$\leq .05$	$\leq .01$	N.S.	$= .06$[b]
Test with brightnesses	$\leq .04$	N.S.	$\leq .01$	$\leq .01$	$\leq .01$	$\leq .01$	$\leq .09$	$= .05$[c]

[a] Each individual Label 1 score is the median RT of *S*'s Label 1 responses to Brightnesses 5–8, minus median RT of his Label 1 responses to Brightnesses 1–4, over median RT of his Label 1 responses to Brightnesses 5–8. Each individual Label 2 score is the median RT of *S*'s Label 2 responses to Brightnesses 9–12, minus median RT of his Label 2 responses to Brightnesses 13–16, over median RT of his Label 2 responses to Brightnesses 9–12.

[b] *P* value for the hypothesis that the three medians lie in this order.

[c] *P* value for the hypothesis that group $\bar{Y}Y$'s median is greater than the others.

performed comparing the three groups during the last 32 training trials, and another was done comparing these groups during the test with brightnesses. As expected, results for neither met conventional significance levels.[11]

Turning now to the latency drop for Label 2 responses from 9–12 to 13–16, we find that for the last 32 training series trials this slope is not significant for the $X\bar{X}$ group, is at the .10 level for the XY group, and is highly significant (.01 level) for Group $\bar{Y}Y$. For the test with brightnesses, the slope again is not significant for Group $X\bar{X}$, but meets the .01 level for both XY and $\bar{Y}Y$. These trends across groups are in keeping with our hypothesis, but the further task remains of determining whether indeed the slope for $\bar{Y}Y$ exceeds that for XY, and that for XY exceeds the $X\bar{X}$ group's slope, as predicted. Since we are predicting that the distributions of the slope scores will fall in a specific order for the three groups, a test by Whitney's extension of the *U* sta-

[11] That results for the test with brightnesses are at the .09 level is probably due to chance, since no comparable trend occurred for the last 32 trials of the training series. We would expect a genuine trend to be reflected in both training and test series, since the trend for Label 2 on the test with brightnesses is, as we shall see later, quite consistent with the Label 2 trend for the last 32 training series trials.

tistic to three groups is called for.[12] Following this procedure, our hypothesis that the slope for Group $\bar{Y}Y$ exceeds that for XY, while XY's slope exceeds $X\bar{X}$'s slope, is significant at the .06 level for the last 32 training series trials. On the test with brightnesses, the number of *S*s in Group $X\bar{X}$ providing reaction to at least one brightness in the range 9–12 and also to at least one in the range 13–16, was so small that the hypothesis of a specific order of degrees of slope for the three groups could not be tested as such. But another version of Whitney's extension of the *U* technique could be applied—one testing the related hypothesis that the slope for Group $\bar{Y}Y$ is greater than that for the other two groups. This prediction was confirmed at the .05 level. On the whole, then, the reaction time data support our expectations quite well.[13]

[12] This test can be used in the present case, whereas it couldn't be applied to Tables 2–5, because the number of intergroup ties is minimal here. In dealing with those few ties that did occur, we followed Jonckheere's (1954) recommendation, which is the most conservative policy possible: namely, assigning the ties in a manner that favors the null hypothesis and is unfavorable to the hypothesis of interest to us. Specifically, ties are excluded from the sum when a large sum is favorable to our hypothesis, and ties are included in the sum when a small sum is favorable to our hypothesis.

[13] It is of interest to consider Cartwright's (1941a, 1941b) work on decision time in the light of our reaction time findings for Label 2 responses. Cartwright suggested the generalization that decision time decreases as one moves away from the boundary and toward the center of a category along some stimulus dimension. Our findings with regard to Label 2 responses suggest, however, that this is not invariably the case, but rather depends, in ambiguous situations, on whether the category in question is defined positively or negatively, and whether its contrast class is defined positively or as the complementary negative. For when the class of interest was defined negatively with a positive contrast class, we found no decrease in decision time as the center of the class was approached from its boundary; when that class was positively defined with a positive contrast class, we found some decrease; and when that class was positively defined and the contrast class was its negative complementary, we found the greatest decrease of all.

Discussion

Our Hypotheses in the Light of the Results

Our hypotheses concerning acceptance limits for placing events in a class may be grouped in two ways. Some concern definition of a class of interest by inclusion or exclusion (namely, Hypotheses 1*a* and 1*b*), while others concern definition of a contrast class by inclusion or exclusion (Hypotheses 2*a* and 2*b*). One may also note that some concern the above differences when the class of interest is ambiguous (Hypotheses 1*a* and 2*a*), while others involve cases where the class of interest is partly certain (Hypotheses 1*b* and 2*b*). We find from Table 7 that Hypotheses 1*a*, 2*a*, and 2*b* tend to be confirmed more strongly than 1*b*. The appropriate conclusions would therefore seem to be these: Whether the class of interest is defined positively or negatively, and whether the contrast class is defined positively or by exclusion of the class of interest, both influence the range of acceptance limits for placing events in the class of interest, but the effect of the latter factor is somewhat greater than that of the former. Further, these effects are somewhat stronger when the class of interest is fully ambiguous than when it is partially certain, although present to a considerable extent even in the latter case.

TABLE 7

Evaluation of the Four Hypotheses in Terms of Some of the Results for Brightnesses 9–16[a]

Confirmations	Hypotheses			
	1*a*	1*b*	2*a*	2*b*
C/N, $P \leq .10$	7/8	2/6	6/8	4/6
C/N, $P \leq .05$	5/8	1/6	4/8	4/6

[a] We have arbitrarily assumed that each chance considered for confirmation of an hypothesis should be given equal weight with any other, and show, for each hypothesis, the number of tests that confirm it (C) divided by the total number of possibilities for its confirmation (N). Results for refraining, "equal" judgments, and indifference points are not considered here since a finding of equality between two groups on those measures may confirm two hypotheses or, alternatively, may confirm neither of them. In the first row, we define confirmations as two-group comparisons yielding *P*'s of .10 or better; in the second row, .05 or better.

Why do these effects occur? Let us consider each of our two basic hypotheses in turn.

On Definition by Inclusion and Exclusion

A reminder is in order concerning the nature of our experimental conditions: The implications as to positively and negatively defined classes for our groups were provided by allusion, rather than by differences in logical structure of the classes. This was done so that the three groups would be equated with regard to the kinds and directness of information they were obtaining, and would differ only in whether a class *seemed* to be defined in its own right, or rather by exclusion of some other class. If the *S*s had been dealing differently with these two kinds of classes in their prior everyday experience (for whatever reason), we would thus expect such differences to transfer to the experimental situation, even though there was no advantage in the experiment proper from treating the two kinds of classes differently. Had there been such an advantage in the present experiment, we would never know if differences found between groups were due solely and specifically to the nature of this advantage or rather were due to the overwhelming influence of class definition by inclusion vs. exclusion as such. Various factors active prior to the time of the experiment could, in turn, cause a preference for positive classes—but our basic question was: Does a preference for inclusion-defined classes exist that is strong enough to operate even when it does not give *S* an advantage in the present experimental situation?

The results of our research indeed suggest that people have a prejudice of this kind in favor of classes defined by inclusion. There are several factors that may underlie this preference.

Disjunctiveness of an exclusion class. To assign an event to an exclusion class is virtually a declaration of ignorance as to how to deal with it, since the number of possible alternative actions excluded by such assignment is usually much smaller than the number of alternatives that remain. If something is not a giraffe, for example, it may be a lion, or a man, or a dandelion, or any one of a myriad of remaining possibilities. The appropriate modes of conduct toward it if it is lion, man, or anything else, are quite different from one another, and we have no basis for choosing one over another. This is the point which seems to underlie Bartlett's (1957) discussion when he notes that transfer of past learning concerning a class is possible when we are presented with new positive exemplars of that class, but not when faced with new negative exemplars of it. A negative instance of a class simply informs you that behavior appropriate to members of that class is not appropriate here, but it does not tell you what constitutes suitable behavior in the present case. The same matter is adumbrated by Hovland and Weiss (1953) when they make the following comment: "The positive instances define a coherent, well-structured combination of characteristics of the concept, while the negative instances are in effect positive instances of what the concept 'is not.' They thus lack an organized, unitary quality, except for the limiting situation where only a single dimension is involved" (p. 181).

Our study, of course, concerns just this limiting situation, since only one dimension—brightness—was used. Although a class that we defined by exclusion was thus just as organized and unitary as one we defined by inclusion, our *S*s dealt with exclusion classes as they might if their definition were disorganized, disjunctive. Such an effect may well be due to transfer from the many cases a person has encountered where exclusion classes' definitions indeed are disjunctive.

Necessity for transforming information when working with an exclusion class. Even if there is actually only one possible alternative class to which an event can belong when one knows that it is not an instance of some other class, the classifier now must usually engage in the added mental operation of assigning the event to that alternative class. ("This is not-X; therefore it is such-and-such.") There is thus one more thing he must do than would be the case if he had been able to place the event in its

appropriate class at the outset. That individuals dislike this extra operation is evident from the preference for "direct test" found by Bruner, Goodnow, and Austin (1956) in their *S*s.

The extent of this preference is well illustrated by comparing particular groups from a study by Bruner, Wallach, and Galanter (in press), and work by Goodnow and Pettigrew (1956). In both cases, *S*'s task was to learn that the correct pattern of pressing a left- and a right-hand button was left-left-right repeated. In the Bruner, Wallach, and Galanter experiment *S* was informed, on any trial, which of the buttons was correct: If the left button were correct on a particular trial, a left-hand light would go on after *S* had pushed either button; while if the right button were correct on that trial, a right-hand light would appear after *S* had pressed either button. In Goodnow and Pettigrew's study, on the other hand, *S* was forced to transform information concerning which button was not correct, on trials in which *S* erred, into knowledge of what button was correct for that trial: If *S* responded correctly on any trial, a poker chip fell out of a central chute; while if he made an error on a trial, no chip appeared. When he pushed the wrong button on a trial, then, *S* had to reason that this trial was not appropriately dealt with by that button (this trial in the sequence was not an instance of the class defined by that button), but rather should be handled by the button on the other side. In contrast, when the *S* in Bruner, Wallach, and Galanter's experiment pushed the wrong button, he immediately knew which button was appropriate for that trial (i.e., he knew to which class—which button—this trial in the sequence should be assigned), since a light went on over the correct button. Positive classes thus were always utilized in the Bruner, Wallach, and Galanter study, since *S* always knew what button was appropriate on that trial. Goodnow and Pettigrew, on the other hand, provided their *S*s with complementary negative classes when they made a mistake—the *E*s told them directly only what button was *not* appropriate on that trial. That Goodnow and Pettigrew's *S*s took about twice as long to learn the left-left-right pattern as Bruner, Wallach, and Galanter's *S*s, indicates how strong an aversion people have against transforming knowledge of what something isn't into knowledge of what it is.

In our experiment, once again, such transformation is not required of *S* when dealing with an exclusion class, since its negative nature is alluded to in the label and nothing more. Placing brightnesses in the exclusion class is no less direct than placing them in the inclusion class. Nevertheless, our *S*s avoided exclusion classes as they might if there were added mental strain involved in their use. Such an effect again may be the result of transfer from the many cases a person has encountered where use of exclusion classes indeed does involve an extra step of reasoning.

Difference in kind of response. Presentation of an instance of an inclusion class requires one to make a particular suitable response of some sort, while presentation of an exemplar of an exclusion class requires one to refrain from making, or withhold, the response that would be appropriate for an instance of the excluded class. Now it may well be that the tendency to make a response is stronger, for the same amount of training, than the tendency to withhold that response; that is, given a certain amount of practice in making or withholding a response, the former tendency may be stronger than the latter. This would lead to a greater inclination to use inclusion classes rather than exclusion classes simply because of the different kinds of responses involved in each.

In our research, again, such a response difference is not obviously involved, since there is no difference in the formal arrangements of button-pushing responses for any of the three conditions: i.e., the $X\bar{X}$ and $\bar{Y}Y$ groups, as well as the XY group, all press one button to indicate placement of a brightness in Class 1, and all press an opposite button to indicate placement of a brightness in Class 2. The *S* indicates that he is using an exclusion class by as active a response as he makes to indicate use of an inclusion class. Further, one cannot argue

that, whereas the verbal labels "X" and "Y" are of equal strength, the label "Yes" is stronger than "No." For our best index of the strength of "Yes" and "No," the Thorndike-Lorge (1944) frequency count, indicates that they are both very strong—and, if anything, the frequency of "No" exceeds that of "Yes." Although no obvious differences in kind of response for inclusion vs. exclusion classes were involved in our experiment, then, the *S*s reacted as they might if such response differences did occur. This once more may be due to transfer from cases where use of inclusion vs. exclusion classes did involve the sort of response differences described.

Innate preference for presence rather than absence of stimulation. We have just noted three reasons why a person might well be expected to learn, during his lifetime, to avoid exclusion classes. Apart from such learning, however, there may also be some innate basis for this preference of inclusion classes. Several theories of motivation—notably those of Freud (1955), Hull (1943), and Miller and Dollard (1941)—have conceived of drives as strong stimuli, and reduction of such stimulation intensity as a preferred state which organisms will strive to achieve. Recently, however, it has been noted that many situations exist in which organisms prefer stimulation increases rather than decreases. For example, Harlow, Harlow, and Meyer (1950) report that monkeys solve mechanical puzzles for the sheer love of doing it; Butler (1953) found that monkeys will solve problems for the "reward" of being able to look out of their cages; Kish (1955) reports that mere onset of illumination will function to strengthen a response; and evidence from the McGill group (Bexton, Heron, & Scott, 1954) indicates that mental life rapidly deteriorates in a human temporarily deprived of stimulus input through the main modalities. Thus, the presence of certain kinds of stimulation may be innately preferred to their absence. If there were a tendency of this sort at the perceptual level, it could well be expected to underlie a preference for positive rather than negative classes at the conceptual level, since a positive class involves the possible presence of its defining properties, while a negative class involves the absence of those properties and the presence of no others in particular.

Several suggestions have been made, then, as to why people prefer positive to negative classes. Any or all of them may actually apply.

On the Relevance of a Contrast Class to a Class of Interest

Our experiment has found that whether a contrast class is defined positively or by exclusion of the class of interest influences the acceptance limits for that latter class. The particular effects exercised by positive vs. negative definition of the contrast class can well be expected from the considerations outlined above, if one assumes that a contrast class will have any influence at all on the class of interest. But it is this latter question that concerns us here. Why should a contrast class influence the acceptance limits for a class of interest in the first place?

Perhaps the answer is somewhat as follows. We have noted elsewhere (Wallach, 1958) how necessary it is for the organism to select some attributes as a basis for establishing equivalence classes and to ignore others, in order to avoid being overwhelmed by the confusing richness of properties that the environment offers the senses. Attention to a contrast class would seem to be one important mechanism for making such selections, since choosing one or another contrast class tends to influence the number and kinds of properties in whose terms we define the class of interest. Further evidence on this point is contained in a recent study by Wallach and Caron (in press).

Although, once again, consideration of the contrast class does not have such an effect in the present experiment since only one dimension—brightness—was used, the demonstrated influence of the contrast class may be due to transfer from the many cases a person has encountered where contrast does serve the function discussed.

Summary and Conclusions

This study was concerned with some of the factors arising from the nature of the classification process itself that may influence our range of acceptance limits for placing events in a category. When it is uncertain whether some event should be placed in a particular class or not—whether this person is a communist, whether that student's grade should be a C, whether there was a speck of light on the telescope's screen—what factors, arising from the way classes are defined, sway our judgment? Two such factors were examined: (*a*) Placing data in a class, we proposed, does not go on in a vacuum, but rather requires that one consider some contrasting class whose nature is used as a basis for deciding how one will specify the class of interest. In particular, we expected people's readiness to assign data to the latter class to be greater when the contrast class was simply defined by exclusion as whatever remained outside the class of interest, than when defined by inclusion as being such-and-such. (*b*) Further, we noted that the class of interest itself may be defined by inclusion or exclusion—that is, in terms of its being such-and-such, or in terms of its not being some other class—and we suggested that people would be more likely to place events in the class of interest when it was defined in the former way.

Seventy university students were distributed among groups designed to test our two hypotheses under each of two conditions of ambiguity of class. It was found that both hypotheses were confirmed: that is, whether the class of interest is defined by inclusion or exclusion, and whether the contrast class is defined by inclusion or by exclusion of the class of interest, both influence the range of acceptance limits for placing events in the class of interest. However, the effect of the latter factor may be somewhat greater than that of the former. Further, these effects were found to be somewhat stronger when the class of interest was highly ambiguous than when it was partially certain, although present to a considerable extent even in the latter case.

Is it significant that these hypotheses were confirmed? Perhaps the best way to answer is to consider their origin. It occurred to us that many matters which psychologists of cognition investigate involve the organism's detecting similarities among events—involve its acting on the assumption that a present event is like something else that happened before. We thought further that such finding of similarities might best be conceptualized as a classification process. If so, judgments as to the range of equivalence—the range of events considered to be similar to one another and hence in a common class—ought to depend on the way the assumed classes are defined. Our hypotheses, in turn, concerned particular differences in such class definitions—differences which were objectively slight in the experiment but which should, on our theory, have large effects. Confirmation of these hypotheses was hence by no means a foregone conclusion, and suggested it may be valuable to explore further implications of the idea that much of cognitive activity involves classification processes.

REFERENCES

BARTLETT, F. C. *Thinking: An experimental and social study.* London: Allen and Unwin, 1957.

BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, **8**, 70–76.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

BRUNER, J. S., WALLACH, M. A., & GALANTER, E. H. The identification of recurrent regularity. *Amer. J. Psychol.*, in press.

BUTLER, R. A. Discrimination learning by rhesus monkeys to visual-exploration motivation. *J. comp. physiol. Psychol.*, 1953, **46**, 95–98.

CARTWRIGHT, D. Decision-time in relation to the differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442. (a)

CARTWRIGHT, D. Relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 174–196. (b)

FREUD, S. Beyond the pleasure principle. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud.* Vol. 18. London: Hogarth Press, 1955. Pp. 7–64.

GOODNOW, J. J., & PETTIGREW, T. F. Some sources of difficulty in solving simple problems. *J. exp. Psychol.*, 1956, **51**, 385–392.

HARLOW, H. F., HARLOW, M. K. & MEYER, D. R. Learning motivated by a manipulation drive. *J. exp. Psychol.*, 1950, **40**, 228–234.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

JONCKHEERE, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

KISH, G. B. Learning when the onset of illumination is used as reinforcing stimulus. *J. comp. physiol. Psychol.*, 1955, **48**, 261–264.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 289–334.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., 1944.

WALLACH, M. A. On psychological similarity. *Psychol. Rev.*, 1958, **65**, 103–116.

WALLACH, M. A., & CARON, A. J. Attribute criteriality and sex-linked conservatism as determinants of psychological similarity. *J. abnorm. soc. Psychol.*, in press.

(Accepted for publication July 16, 1958)